Data science for mental health – a UK perspective on a global challenge

Andrew M McIntosh[*,1], Robert Stewart[2], Ann John[3], Daniel J Smith[4], Katrina Davis[2], Cathie Sudlow[1], Aiden Corvin[5], Kristin K Nicodemus[10], David Kingdon[6], Lamiece Hassan[7], Matthew Hotopf[2], Stephen M Lawrie[1], Tom C Russ[1], John R Geddes[8], Miranda Wolpert[9], Eva Wölbert[11], David J Porteous[10], and the MQ Data Science Group[11]

Affiliations

1 Division of Psychiatry, University of Edinburgh, Edinburgh, UK
2 King's College London (Institute of Psychiatry, Psychology and Neuroscience), London, UK
3 Swansea University Medical School, Swansea University, Swansea, UK
4 Institute of Health and Wellbeing, University of Glasgow, Glasgow, UK
5 Department of Psychiatry & Psychosis Research Group, Trinity College Dublin, Dublin, Ireland
6 Faculty of Medicine, University of Southampton, Southampton, UK
7 Health eResearch Centre, University of Manchester, Manchester, UK
8 Department of Psychiatry, University of Oxford, Oxford UK
9 Child Outcomes Research Consortium (CORC) and Evidence Based practice Unit, University College London, and Anna Freud Centre, London, UK
10 Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK
11 MQ Transforming mental health, London, UK

Corresponding author*

Abstract	111 words
Main text	2980 words
Tables	0
Figures	2

Abstract

Data science extracts new knowledge from high dimensional datasets through computer science and statistics. Mental health research, diagnosis and treatment can benefit from data science using consented cohort studies, genomics, routine healthcare and administrative data. The UK is well placed to trial these approaches through well annotated and NHS-linked data science projects, such as UK Biobank, Generation Scotland and the Clinical Record Interactive Search (CRIS) programme. Data science has great potential as a low cost, high return catalyst for how mental health problems may be better recognised, understood, supported and outcomes improved. Lessons learnt from such studies have the potential for global reach in terms of both their output and impact.

WHAT IS DATA SCIENCE?

Data science is the extraction of knowledge from high-volume data, using skills in computing science, statistics and the specialist domain knowledge of experts[1]. Data science pervades global business and modern living and can partner technical revolutions, such as medical genomics and imaging, to revolutionise the monitoring, diagnosis, treatment and prevention of disease. These ideals are implicit in 'stratified medicine' and 'precision medicine'. The case for data science is often made for cancer, heart and infectious diseases. Here we argue for the enormous potential for data science to transform mental health research and clinical practice worldwide. International collaboration will be necessary for maximum reach and impact. We review the available resources, barriers and opportunities from a UK perspective before setting out how the full potential of data science could be realised on a global scale.

Figure 1: What is data science?

Insert Figure 1 here

WHY MENTAL HEALTH AND WHY NOW?

Mental disorders are arguably the greatest 'hidden' burden of ill health, with substantial long-term impacts on individuals, carers and society[2]. People with these conditions are often socially excluded[3] and less likely to participate in research studies or remain in follow up[4-6]. Complexities around defining diagnoses present particular challenges for mental health research. Richly annotated, longitudinal datasets matched to data science analytics offer an unprecedented opportunity for more robust diagnostics, and also the prediction of outcome, treatment response, and patient preferences to inform interventions[7]. It may also provide more effective targeting of recruitment to observational and interventional studies. Such data are large in size and dimensions and require the application of advanced analytics, such as machine learning, where more conventional techniques are less computationally tractable.

A key issue in data science is the description of data types that are the most informative, readily available and efficiently captured. Generic data types include electronic health and prescribing records, education, welfare, socio-demographic, laboratory and real world monitoring through wearable devices and environmental sensors. More specific data might include genomic data, *in vivo* brain imaging and cognitive traits. Important challenges include shortcomings in dataset completeness and linkage potential, as well as acceptability to patients and the wider public, given the perceived sensitivity of mental health data. It is also important to consider the types of information that can create new ways of classifying mental health and illness, and be universally applied beyond the 'perfect world' discovery setting.

WHAT UK RESOURCES MIGHT HELP PIONEER THIS APPROACH?

1. Population cohorts

There are several UK population cohorts with enhanced clinical, biological and social datasets linked to routinely collected electronic data. We provide details of UK Biobank (www.ukbiobank.ac.uk) here.

UK Biobank

UK Biobank is a cohort study of 0.5 Million individuals aged between 37 and 73 years recruited between 2006 and 2010. Participants completed a touch-screen questionnaire, underwent an interview, and participated in several assessments including measures of depressive symptoms, distress, cognition and alcohol and cigarette use. In addition, linkages have been made to National Health Service (NHS) healthcare episode data, and a number of biological measures have been taken, including DNA for whole-genome genotyping. An initial pilot medical imaging study includes unprocessed brain structure, function and connectivity data in over 5,000 participants, which is in the process of being extended to 100,000 individuals. Further longitudinal and outcome assessments include repeat cognitive testing and actigraphy. Lifetime history of mental illness will be assessed in greater depth with a web-based questionnaire. UK Biobank thus brings unprecedented deep and broad phenotyping to mental health research[8].

There are several other UK population cohorts with deeply-phenotyped participants and the potential for record linkage to routine healthcare and administrative data. Notable examples include The 'Generation Scotland: Scottish Family Health Study' (GS:SFHS)[9,10], a family and population based study located in Scotland with near complete record linkage and the Avon Longitudinal Study of Parents and Children[11], a UK-based cohort study with data available from before birth to more than 20 years follow up. Further information on these and other studies is provided in the online appendix.

2. Domain specific cohorts linked to routinely collected data

In contrast to population based research cohorts, several UK resources are focused on Mental Health and routinely collected clinical data from the NHS, the UK's comprehensive healthcare provider. These data may be more representative of the general population and provide a framework for implementation.

The The National Centre for Mental Health and SAIL Databank

The National Centre for Mental Health (NCMH) was established in Wales in 2011 and partnered to the MRC Centre for Neuropsychiatric Genetic and Genomics. The NCMH recruits participants with mental disorders to the NCMH cohort, currently at over 6000 individuals, who are willing to participate in research and be recontacted. Clinical data (e.g. demographic, routine secondary care, enhanced clinical, neuropsychological, imaging) and biological samples are collected creating a platform and infrastructure for mental health research into the causes and treatment of mental illness and learning disability (www.ncmh.info). In 2015 the Farr Institute partnered with the NCMH allowing for linkage of the cohort to routine data nested within prevalent diagnostic electronic cohorts within the Secure Anonymised Information Linkage (SAIL) databank (www.saildatabank.com)[12,13]. The SAIL Databank is a whole population based research data repository holding over 2Bn anonymised health records, from ~3.5M patients, from primary care, hospitals, child health, education, cause-specific mortality, deprivation and urbanicity. Participants can be tracked across health and social care settings, whilst protecting privacy in accordance with relevant legislation using a split file approach[12,13]. This is the first time genomic data has been linked to the SAIL databank[14] allowing researchers to

address questions on the impact of genetic, environmental and health factors including modifiable lifestyle factors on clinically important outcomes.

3. *Electronic health record derived cohorts and the Farr Institute*

The increasing use of electronic health records is creating databases unparalleled both in sample size and in the depth of information contained. The use of these data for research is encouraged by policy[15,16] and subject to necessary technical and ethical considerations[17-19].

An important distinction is made between structured information and unstructured text – the former being simpler to analyse, albeit that clinical uncertainties are often poorly coded[20-23]. Here, text mining may be employed alongside structured information to better define groups[24,25]. Structured information on patients requiring specialist care has been collected systematically by the NHS since 1981 through Hospital Episode Statistics in England, the Scottish Morbidity Record and Patient Episode Data for Wales. These are available to researchers as linked-data and are published in open-access aggregated form[26,27], along with primary care data[28],[1]. Despite concerns about the speed and accuracy of these electronic data[29,30], these resources may prove valuable for measuring real-world outcomes and assessing their mediators and predictors.

In 2013 electronic medical record linkage was given further impetus by the founding of the UK Farr Institute for Health Informatics Research (see online appendix). It has the aim of harnessing health data for patient and pubic benefit by facilitating the safe and secure use of electronic patient records and other population based data sets.

The Clinical Record Interactive Search (CRIS) application

The CRIS application was developed at the South London and Maudsley NHS Foundation Trust in 2007 as a means of rendering the large volumes electronic mental health record data available for research[31,32]. CRIS accesses mental health case records from around 260,000 patients within a south London geographic catchment of approximately 1.2m residents; replications of CRIS have recently become operational elsewhere in London, Oxford and Cambridge.

Key to the development include data structuring and de-identification pipelines and also a wider data security and governance model which has been patient-led from the outset[33]. Research applications have included searches to help identify and characterise rare scenarios for further investigation[34,35], and data linkage projects to characterise physical health outcomes[36,37]. Recent enhancements include the development of natural language processing applications to derive structured information from the text fields present in the electronic mental health record. These include recorded diagnoses, cognitive test scores, pharmacotherapy and symptom profiles[38-42].

The Child Outcomes Research Consortium approach is also a flagship electronic record UK project and is based around the outcomes of children and adolescents seen in specialist mental health services[43]. Further details are provided in the online appendix.

Linkage to 'real-time' health data and wearable devices

Companies such as Apple (Healthkit and Researchkit) and Google (Alphabet) are developing health based applications and wearable devices, as part of a wider array of environmental sensors, 'The Internet of Things', and health application developer toolkits. The potential to capture new sources of relevant 'real-time' and longitudinal health data (e.g. mood, diet, activity and sleep patterns), matched to physiological measures (e.g. of heart rate, blood glucose, cortisol) is potentially transformative and pervasive at low cost and independent of conventional healthcare provision. A good, early example of such an initiative in psychiatry is Truecolours (https://oxfordhealth.truecolours.nhs.uk/www/en/), a platform developed to capture continuous patient-generated data with the required usability and acceptability to permit reliable longitudinal follow-up. It is also notable that this technology is currently being piloted as a supplement to routine healthcare.

PUBLIC TRUST AND CLINICAL GOVERNANCE

Whilst UK data science resources represent major opportunities for research and health service improvement, they demand public support, public trust and transparent governance arrangements. The MRC Farr Institute, the European Data in Health Research Alliance (datasaveslives.eu) and Patients4Data

(patients4data.co.uk) are all promoting the importance of data sharing for research and healthcare impact whilst acknowledging the potential risks of inaccurately recorded information and data breaches.

Attitudes research suggests that mental health data are among the most personal and sensitive[44,45]. There are diverse reasons why people might be reluctant or unwilling to consent to the use of their data for mental health research[46,47]. Studies indicate, encouragingly, that a majority of mental health service users agree to the use of their health records for research – particularly when efforts to engage in on-going communication about their use and potential benefits are made[32,48]. It is important to reflect how cancer research has largely dispelled the past stigma of a cancer diagnosis: can modern day research, driven by data science, do the same for mental health? We think so, by reframing and redefining the causes and by reshaping and revitalising effective interventions.

Safe and transparent models of governance for re-use of mental health data are essential for maintaining public trust. Systems have been developed that protect privacy and, in future, innovations such as dynamic models of consent[49] may also allow the public further control over their data. The recently established Farr Institute (see appendix) includes a programme of public engagement with a focus on the safe and transparent use of patient and research data.

The 'Scottish Model'

The 'Scottish Model' is a useful illustrative example of how data science and record linkage can be conducted at scale and in a trusted environment. Like many Scandinavian countries, Scotland hosts excellent administrative and healthcare data resources. The NHS Community Health Index (CHI) - a unique personal identifier for 99% of the population, has greatly enabled pseudonymised linkage between health and administrative data (Figure 2).

Figure 2. National level data resources in Scotland

Insert Figure 2 here

Arguably, the other key to unlocking the benefits of routinely collected data in Scotland has been the presence of good research governance procedures and proactive engagement with the public to drive forward health informatics research. Public input into reviewing grant applications is standard practice, and includes providing later lay research summaries and wider dissemination in addition to public consultation and outreach. Consultation work suggests that the public supports the use of administrative and health data in research, provided there is adequate data security and access is limited to personnel conducting research for public benefit. The public appears more supportive of academic and clinical research than work conducted by commercial organisations[44,50].

All data outputs are scrutinised to ensure they do not identify individuals or breach privacy before being released. Open access summaries are published online as a condition of all research. Support to researchers throughout this process is provided by an eData Research and Innovation Service[51]. The key elements of the Scottish model are illustrated within Figure 2 (adapted from a previous publication[51]).

TRAINING, RESOURCE AND CAPACITY IMPLICATIONS

The availability and development of excellent resources for data science alongside robust governance procedures are necessary prerequisites for good data science. We would argue that there are also specific technological and skills challenges to be overcome and that fulfilling the promise of data science will involve international collaboration spanning high and low income countries.

1. Technological resource

The capacity of data storage and access, and the personnel to collect and analyse data are rate-limiting steps in the ongoing development of data science. Routinely-collected 'administrative' and health data tend to be centrally financed by government but have limited phenotypic coverage and have, until recently, been used mainly for planning. More detailed phenotyping is possible in routine clinical data, such as CRIS in London and PsyCIS in Glasgow[52], and large scale genetic, '-omics' and neuroimaging studies generate huge volumes of data that

pose tractable data storage issues. The combination of these datasets is very challenging and requires data harmonisation and for compatibility issues to be addressed.

Databases need to gather and hold data, and enable users to search for and access data of interest to them. Data sharing agreements and how to facilitate collaboration and innovation are key issues for data scientists. In practice, data generation projects are deciding on a case by case basis what they they will offer to centralised depositories without offering a coordinated solution for how that data will be linked to other sources. Centralised databases can make themselves more attractive to data depositors by offering managed data access and trusted analysis environments. The Global Alliance for Genomics and Health (https://genomicsandhealth.org) is an international example that brings together different health sectors and regions worldwide, to catalyse the sharing of methods to harmonise data approaches across diverse datasets.

2. Skills resource

Identifying, training and fostering a generation of clinically-informed data scientists from a wide range of backgrounds must be a top priority. This requires multidisciplinary training programmes, which expose scientists, informaticians and statisticians to commonly used clinical data, diagnoses and treatments, as well as a range of relevant methodological approaches. Data scientists will usually need further postgraduate training in statistics and computational methods. Trainees will need to be familiar with ethical and regulatory requirements as well as prepared to become familiar with the diverse ways in which health data are recorded and stored. Given the diversity of resources and methodologies, a variety of approaches seems inevitable. Particular care and attention to the career structure of data scientists will be needed to nurture early-career researchers and ensure that expensively acquired expertise is not lost after training. A spectrum of skills and disciplines needs to be present in a data science team and its leadership as well as a common understanding of the need for complementary expertise. As data science evolves in fields such as engineering and finance, there will be opportunities to learn from their experience.

3. National and international collaboration

In order to achieve maximum reach and impact, there is a need to develop and maintain international and interdisciplinary databases and the networks to support their efficient use. There are many examples of this process working well in areas such as genomics[53] and brain imaging[54], where international consortia have brought together databases of unparalleled size and scope. There are particular challenges in expanding these initiatives to low and middle-income countries where the infrastructure may be more limited and low cost methods of data collection and storage will be needed. Clinical information from paid and public health providers may also come with differing governance frameworks and commercial interests, but overcoming these barriers will prove beneficial for all parties.

There is also much work to be done in standardising assessments, outcome measures and terminology within, let alone between, nations. UK and international research charities such as MQ, the Wellcome Trust and publicly funded research councils have an important role to play in matching researchers and their research questions to datasets spanning multiple subject domains and countries. Routine health record data with detailed mental health coverage are stored in parts of the UK, Australia as well as the exemplary Scandinavian systems. Some projects, like UK Biobank, encourage external data analysis even as data are being collected, whereas others will not be openly shared until the original funder-approved aims have been met. Subject to regulatory approvals, it is desirable that systems should be put in place to facilitate the incorporation of data from time-limited projects as soon as practicable. Intellectual property and resource considerations may make this challenging. Fostering collaborations, developing safe havens to facilitate joint working and convening advisory groups with wide representation will help enhance complementarity across projects and data collections.

OUR VISION OF THE FUTURE

Against a backdrop of no fundamentally new pharmacologic treatment in the past 60 years and a progressive pharmaceutical industry withdrawal from mental health Research and Development, an alternative course is essential. Mental

health remains the leading area of unmet medical need in the developed world, and is rapidly acquiring the same status in the developing world.

Combining large healthcare and administrative datasets with real-time monitoring, laboratory, genomic and imaging data could achieve a step change in the way healthcare is provided and research is organised. In our opinion, data science will greatly enhance our ability to conduct discovery science, epidemiological studies, personalised medicine and plan services. Without the better understanding of mental health problems that will come with use of Big Data, longer term visions for self-management, better treatments and learning health systems will not be possible. It is thus vital that current initiatives in data science recognise and support this need.

Contributors

The manuscript was critically revised by:

Gerard Leavey, The Bamford Centre for Mental Health and Wellbeing, University of Ulster, United Kingdom. Graham Moon, Geography and Environment, University of Southampton, United Kingdom. Rosie Cornish, School of Social and Community Medicine, University of Bristol, United Kingdom. Tamsin Ford, University of Exeter Medical School, Exeter, United Kingdom. Gary Donohoe, Center for Neuroimaging and Cognitive Genomics (NICOG), School of Psychology, NUI Galway, United Kingdom. Rudolf Cardinal, Department of Psychiatry, University of Cambridge, United Kingdom. Zina Ibrahim, Department of Social Genetic & Developmental Psychiatry Kings College London, United Kingdom. Margaret Maxwell, NMAHP Research Unit, School of Health Sciences, University of Stirling, Stirling, United Kingdom. Nadine Dougall, NMAHP Research Unit, School of Health Sciences, University of Stirling, Stirling, United Kingdom. Felicity Callard, PhD, Department of Geography, University of Durham, United Kingdom. David McDaid, Personal Social Services Research Unit, London School of Economics and Political Science, London WC2A 2AE, United Kingdom

Conflicts of interest

Andrew M McIntosh has received research funding from Pfizer, Janssen and Eli Lilly

Robert Stewart has received research funding from Pfizer, Lundbeck, Roche, Janssen and GSK

Ann John reports no conflicts of interest

Daniel J Smith reports no conflicts of interest

Katrina Davis reports no conflicts of interest

Cathie Sudlow reports no conflicts of interest

Aiden Corvin reports no conflicts of interest

Kristin K Nicodemus reports no conflicts of interest

David Kingdon reports no conflicts of interest

Lamice Hassan reports no conflicts of interest

Matthew Hotopf reports no conflicts of interest

Stephen M Lawrie has received grants and personal fees from Roche, grants from Pfizer and Abbvie, and personal fees from Janssen and Sunovion.

Tom C Russ reports no conflicts of interest

John R Geddes reports no conflicts of interest

Miranda Wolpert reports no conflicts of interest

Eva Wölbert reports no conflicts of interest

David J Porteous reports no conflicts of interest

Legend to Figure 1 What is data science:

Figure showing the components of data science.

Legend to Figure 2: The 'Scottish' Model

Figure shows the linkable data sources available in Scotland, whose linkage is facilitated by the unique CHI number. Administrative data is shown separately from NHS data in the lower panel.

References

1. Dhar V. Data Science and Prediction. *Commun Acm* 2013; **56**(12): 64-73.

2. Whiteford HA, Degenhardt L, Rehm J, et al. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *Lancet* 2013; **382**(9904): 1575-86.

3. Barr B, Kinderman P, Whitehead M. Trends in mental health inequalities in England during a period of recession, austerity and welfare reform 2004 to 2013. *Soc Sci Med* 2015; **147**: 324-31.

4. van Heuvelen MJG, Hochstenbach JBM, Brouwer WH, et al. Differences between participants and non-participants in an RCT on physical activity and psychological interventions for older persons. *Aging Clinical and Experimental Research* 2005; **17**(3): 236-45.

5. Rogers A, Harris T, Victor C, et al. Which older people decline participation in a primary care trial of physical activity and why: insights from a mixed methods approach. *BMC Geriatr* 2014; **14**.

6. Goldberg M, Chastang JF, Zins M, Niedhammer I, Leclerc A. Health problems were the strongest predictors of attrition during follow-up of the GAZEL cohort. *Journal of clinical epidemiology* 2006; **59**(11): 1213-21.

7. Torous J, Baker JT. Why Psychiatry Needs Data Science and Data Science Needs Psychiatry: Connecting With Technology. *JAMA Psychiatry* 2016; **73**(1): 3-4.

8. Smith DJ, Nicholl BI, Cullen B, et al. Prevalence and characteristics of probable major depression and bipolar disorder within UK biobank: cross-sectional study of 172,751 participants. *PLoS One* 2013; **8**(11): e75362.

9. Smith BH, Campbell A, Linksted P, et al. Cohort profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. *International journal of epidemiology* 2012.

10. Smith BH, Campbell H, Blackwood D, et al. Generation Scotland: the Scottish Family Health Study; a new resource for researching genes and heritability. *BMC Med Genet* 2006; **7**: 74.

11. Fraser A, Macdonald-Wallis C, Tilling K, et al. Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int J Epidemiol* 2013; **42**(1): 97-110.

12. Ford DV, Jones KH, Verplancke JP, et al. The SAIL Databank: building a national architecture for e-health research and evaluation. *BMC Health Serv Res* 2009; **9**: 157.

13. Lyons RA, Jones KH, John G, et al. The SAIL databank: linking multiple health and social care datasets. *BMC medical informatics and decision making* 2009; **9**: 3.

14. Lloyd K, McGregor J, John A, et al. A national population-based e-cohort of people with psychosis (PsyCymru) linking prospectively ascertained phenotypically rich and genetic data to routinely collected records: overview, recruitment and linkage. *Schizophrenia research* 2015; **166**(1-3): 131-6.

15. Department of Health. Personalised health and care 2020: Using data and Technology to Transform Outcomes for Patients and Citizens. London: HM Givernment, 2014.

16. Clarke A, Adamson J, Sheard L, Cairns P, Watt I, Wright J. Implementing electronic patient record systems (EPRs) into England's acute, mental health and community care trusts: a mixed methods study. *BMC medical informatics and decision making* 2015; **15**: 85.

17. Coorevits P, Sundgren M, Klein GO, et al. Electronic health records: new opportunities for clinical research. *J Intern Med* 2013; **274**(6): 547-60.

18. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. *Nature reviews Genetics* 2012; **13**(6): 395-405.

19. Nuffield Council on Bioethics. The collection, linking and use of data in biomedical research and health care: ethical issues, 2015.

20. Morrison Z, Fernando B, Kalra D, Cresswell K, Sheikh A. National evaluation of the benefits and risks of greater structuring and coding of the electronic health record: exploratory qualitative investigation. *J Am Med Inform Assoc* 2014; **21**(3): 492-500.

21. Delaney BC, Peterson KA, Speedie S, Taweel A, Arvanitis TN, Hobbs FD. Envisioning a learning health care system: the electronic primary care research network, a case study. *Ann Fam Med* 2012; **10**(1): 54-9.

22. Bernat JL. Ethical and quality pitfalls in electronic health records. *Neurology* 2013; **81**(17): 1558.

23. Whooley O. Diagnostic ambivalence: psychiatric workarounds and the Diagnostic and Statistical Manual of Mental Disorders. *Sociol Health Illn* 2010; **32**(3): 452-69.

24. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *J Am Med Inform Assoc* 2013; **20**(1): 144-51.

25. Denny JC. Chapter 13: Mining electronic health records in the genomics era. *PLoS computational biology* 2012; **8**(12): e1002823.

26. Sinha S, Peach G, Poloniecki JD, Thompson MM, Holt PJ. Studies using English administrative data (Hospital Episode Statistics) to assess health-care outcomes-systematic review and recommendations for reporting. *European journal of public health* 2013; **23**(1): 86-92.

27. Health & Social Care Information Centre. Users and Uses of Hospital Episode Statistics, 2012.

28. Heath & Social Care Information Centre. Supporting open data and transparency, 2015.

29. RSA Open Public Services Network. Exploring how available NHS data can be used to show the inequality gap in mental healthcare, 2015.

30. CAPITA. The quality of clinical coding in the NHS, 2014.

31. Perera G, Broadbent M, Callard F, et al. Cohort profile of the South London and Maudsley NHS Foundation Trust Biomedical Research Centre (SLaM BRC) Case Register: current status and recent enhancement of an Electronic Mental Health Record-derived data resource. *BMJ open* 2016; **6**(3): e008721.

32. Stewart R, Soremekun M, Perera G, et al. The South London and Maudsley NHS Foundation Trust Biomedical Research Centre (SLAM BRC) case register: development and descriptive data. *BMC Psychiatry* 2009; **9**: 51.

33. Fernandes AC, Cloete D, Broadbent MT, et al. Development and evaluation of a de-identification procedure for a case register sourced from mental health electronic records. *BMC medical informatics and decision making* 2013; **13**: 71.

34. Su YP, Chang CK, Hayes RD, et al. Retrospective chart review on exposure to psychotropic medications associated with neuroleptic malignant syndrome. *Acta Psychiatr Scand* 2014; **130**(1): 52-60.

35. Oram S, Khondoker M, Abas M, Broadbent M, Howard LM. Characteristics of trafficked adults and children with severe mental illness: a historical cohort study. *Lancet Psychiatry* 2015; **2**(12): 1084-91.

36. Chang CK, Hayes RD, Perera G, et al. Life expectancy at birth for people with serious mental illness and other major disorders from a secondary mental health care case register in London. *PLoS One* 2011; **6**(5): e19590.

37. Chang CK, Hayes RD, Broadbent MT, et al. A cohort study on mental disorders, stage of cancer at diagnosis and subsequent survival. *BMJ open* 2014; **4**(1): e004295.

38. Patel R, Jayatilleke N, Broadbent M, et al. Negative symptoms in schizophrenia: a study in a large clinical sample of patients using a novel automated method. *BMJ open* 2015; **5**(9): e007619.

39. Patel R, Lloyd T, Jackson R, et al. Mood instability is a common feature of mental health disorders and is associated with poor clinical outcomes. *BMJ open* 2015; **5**(5): e007504.

40. Perera G, Khondoker M, Broadbent M, Breen G, Stewart R. Factors associated with response to acetylcholinesterase inhibition in dementia: a cohort study from a secondary mental health care case register in london. *PLoS One* 2014; **9**(11): e109484.

41. Kadra G, Stewart R, Shetty H, et al. Extracting antipsychotic polypharmacy data from electronic health records: developing and evaluating a novel process. *BMC Psychiatry* 2015; **15**: 166.

42. Hayes RD, Downs J, Chang CK, et al. The effect of clozapine on premature mortality: an assessment of clinical monitoring and other potential confounders. *Schizophr Bull* 2015; **41**(3): 644-55.

43. Fleming I, Jones M, Bradley J, Wolpert M. Learning from a Learning Collaboration: The CORC Approach to Combining Research, Evaluation and Practice in Child Mental Health. *Administration and Policy in Mental Health and Mental Health Services Research* 2014; **43**(3): 297-301.

44. Wellcome Trust. Qualitative Research into Public Attitudes to Personal Data and Linking Personal Data, 2013.

45. Taylor MJ, Taylor N. Health research access to personal confidential data in England and Wales: assessing any gap in public attitude between preferable and acceptable models of consent. *Life Sci Soc Policy* 2014; **10**: 15.

46. Ridgeway JL, Han LC, Olson JE, et al. Potential Bias in the Bank: What Distinguishes Refusers, Nonresponders and Participants in a Clinic-Based Biobank? *Public Health Genomics* 2013; **16**(3): 118-26.

47. Papoulias C, Robotham D, Drake G, Rose D, Wykes T. Staff and service users' views on a 'Consent for Contact' research register within psychosis services: a qualitative study. *Bmc Psychiatry* 2014; **14**.

48. Callard F, Broadbent M, Denis M, et al. Developing a new model for patient recruitment in mental health services: a cohort study using Electronic Health Records. *BMJ open* 2014; **4**(12).

49. Williams H, Spencer K, Sanders C, et al. Dynamic Consent: A Possible Solution to Improve Patient Confidence and Trust in How Electronic Patient Records Are Used in Medical Research. *Jmir Med Inf* 2015; **3**(1).

50. Willison DJ, Steeves V, Charles C, et al. Consent for use of personal information for health research: Do people with potentially stigmatizing health conditions and the general public differ in their opinions? *Bmc Med Ethics* 2009; **10**.

51. Pavis S, Morris AD. Unleashing the power of administrative health data: the Scottish model. *Public Health Res Pr* 2015; **25**(4).

52. Martin DJ, Park J, Langan J, Connolly M, Smith DJ, Taylor M. Socioeconomic status and prescribing for schizophrenia: analysis of 3200 cases from the Glasgow Psychosis Clinical Information System (PsyCIS). *Psychiatr Bull (2014)* 2014; **38**(2): 54-7.

53. O'Donovan MC. What have we learned from the Psychiatric Genomics Consortium. *World Psychiatry* 2015; **14**(3): 291-3.

54. Thompson PM, Stein JL, Medland SE, et al. The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain Imaging Behav* 2014; **8**(2): 153-82.